



EMPIRE: a highly parallel semiempirical molecular orbital program: 3: Born-Oppenheimer molecular dynamics

Johannes T. Margraf¹ · Matthias Hennemann² · Timothy Clark²

Received: 5 December 2019 / Accepted: 13 January 2020 / Published online: 3 February 2020
© The Author(s) 2020

Abstract

Direct NDDO-based Born-Oppenheimer molecular dynamics (MD) have been implemented in the semiempirical molecular orbital program EMPIRE. Fully quantum mechanical MD simulations on unprecedented time and length scales are possible, since the calculation of self-consistent wavefunctions and gradients is performed in a massively parallel manner. MD simulations can be performed in the NVE and NVT ensembles, using either deterministic (Berendsen) or stochastic (Langevin) thermostats. Furthermore, dynamics for condensed-phase systems can be performed under periodic boundary conditions. We show three exemplary applications: the dynamics of molecular reorganization upon ionization, long timescale dynamics of an endohedral fullerene, and calculation of the vibrational spectrum of a nanoparticle consisting of more than eight hundred atoms.

Keywords Massively parallel · Born-Oppenheimer molecular dynamics · Semiempirical MO theory · EMPIRE

Introduction

EMPIRE is a highly parallel semiempirical molecular orbital program [1, 2]. It uses MNDO-like neglect of differential diatomic overlap (NDDO) [3, 4] methods to calculate the electronic structure, geometry, and properties of molecules and periodic systems [5–12]. The main incentive behind combining these very efficient methods with massively parallel programming is that this allows the full quantum mechanical simulation of systems consisting of tens of thousands of atoms. Such system sizes are quickly reached in realistic simulations of biological systems, nanostructures or condensed, amorphous structures such as liquids or organic solids [13].

Thanks to specialized codes and increased computer power, quantum mechanics-based MD simulations (particularly

using DFT) have become common [14]. Their high accuracy makes DFT-based MD simulations very attractive, but the cost of such calculations restricts their use to small systems and short simulation times. This can be problematic because poor sampling undermines the quality of the results. At the other end of the spectrum, classical MD simulations allow very long MD trajectories to be calculated for millions of atoms. Here, the quality of the simulation depends strongly on the quality of the empirical potential energy function, which can be difficult to determine. Additionally, many important effects such as polarization, charge transfer, and bond breaking cannot be included routinely because the electrons of the system are effectively coarse grained in the force field.

NDDO- or tight-binding density-functional theory (TBDFT) [15]–based semiempirical methods present an ideal compromise between these extremes; they have therefore been used extensively in quantum mechanical molecular dynamics studies. Historically, the use of these methods is strongly connected to mixed QM/MM simulations [16], where only a part of the system is treated quantum mechanically (e.g. the reaction center of a protein), while the rest is treated classically. Thus, the popular classical MD codes CHARMM and AMBER actually include specific modules for semiempirical calculations [17, 18]. Particularly modern dispersion-corrected semiempirical methods like PM6-DH2 [19, 20] and OMx-D3 [21] have emerged as economical and accurate methods in QM/MM simulations of biological systems [19, 22–25].

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00894-020-4293-z>) contains supplementary material, which is available to authorized users.

✉ Timothy Clark
Tim.Clark@fau.de

¹ Chair of Theoretical Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85747 Garching, Germany

² Computer-Chemie-Centrum, Department of Chemistry and Pharmacy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany

However, the full potential of the semiempirical methods is realized in fully quantum mechanical MD simulations, where they can tackle systems that are simply too large for DFT or ab initio methods. Grimme recently used OM2-D3 MD simulations to predict the fragmentation of molecules in mass spectrometers accurately [19, 26]. Bartlett et al. used a system-specific “transfer Hamiltonian” to study the breaking of a silica nanorod under tensile stress at near coupled-cluster accuracy [27–30]. Another interesting aspect of is the use of semiempirical MD to study the dynamics of excited states [31, 32].

Despite the efficiency of semiempirical techniques, moving to system sizes beyond a few hundred atoms has proven to be difficult with NDDO or TB-DFT standard implementations. As a consequence, several linear-scaling schemes have been developed, based on the partial localization of the wavefunction [33, 34]. While these schemes allow impressively large calculations on a simple desktop computer, the localization is not necessarily a valid approximation for certain systems. This is particularly true for conjugated systems or any system with a low HOMO-LUMO gap, including zwitterionic proteins [35].

The main idea behind EMPIRE is to be able to simulate unprecedented system sizes fully quantum mechanically, without resorting to linear-scaling algorithms. This is achieved by using massively parallel computer clusters. This philosophy can easily be extended to MD simulations, since the rate-determining step is the already parallel SCF procedure. In this software report, we discuss the implementation and performance of Born-Oppenheimer MD (BOMD) in EMPIRE, and show some examples of successful applications.

Theory

The Born-Oppenheimer approximation [36] states that the dynamics of electrons and nuclei in a molecule can be decoupled due to the large mass difference between them. In the context of MD simulations, this means that electrons in a molecule instantaneously adapt to changes in the configuration of the nuclei, i.e. the electronic wavefunction remains exactly in the variational ground state. In a BOMD simulation, a full SCF calculation of the wavefunction is therefore conducted at each timestep. This procedure requires more computation time per timestep than the Car-Parinello scheme [37], in which the electronic degrees of freedom are propagated together with the atom positions in the MD, but allows a larger time increment to be used. As a result, the computational demand of the two approaches is comparable.

The forces acting on each atom are calculated at every timestep, t , based on the current wavefunction and molecular geometry. Using this information, the geometry at the next timestep $t + \Delta t$ is determined via numerical integration of Newton’s equations of motion, using the Velocity-

Verlet algorithm [38]. In particular, for each atom A ,

$$\mathbf{r}_A(t + \Delta t) = \mathbf{r}_A(t) + \mathbf{v}_A(t)\Delta t + \frac{\mathbf{a}_A(t)\Delta t^2}{2}, \text{ where} \quad (1)$$

$$\begin{aligned} \mathbf{a}_A(t + \Delta t) &= \frac{\mathbf{F}_A(t + \Delta t)}{m_A} \text{ and} \\ \mathbf{v}_A(t + \Delta t) &= \mathbf{v}_A(t) + \frac{\Delta t[\mathbf{a}_A(t) + \mathbf{a}_A(t + \Delta t)]}{2} \end{aligned} \quad (2)$$

where \mathbf{r}_A , \mathbf{a}_A , \mathbf{v}_A are the position, acceleration, and velocity vectors of atom A , and \mathbf{F}_A and m_A are its force vector and mass.

The Velocity-Verlet algorithm performs dynamics in the microcanonical ensemble (also called the NVE ensemble, since the number of particles, volume, and total energy are conserved). In many cases, however, simulations at constant temperature (i.e., NVT, the canonical ensemble) are desirable, because they provide a more realistic model of common experimental conditions. This is achieved by modifying the equations of motion using a thermostat.

According to the Maxwell-Boltzmann equation, the temperature is proportional to the average kinetic energy of the particles in a system. Maintaining constant temperature therefore requires a mechanism that allows the transfer of kinetic energy to and from an external heat bath whenever the average temperature deviates from the desired value. We can calculate the instantaneous temperature at a given point in time using the kinetic energy of all atoms in the system:

$$T = \frac{2E_{kin}}{N_f k_B} \quad (3)$$

where N_f is the number of degrees of freedom in the system. The average of the instantaneous temperature T corresponds to the thermodynamic temperature T_0 . Please note that sampling the canonical (NVT) ensemble does not mean that the instantaneous temperature remains constant. In fact, fluctuations of the instantaneous temperature are expected, particularly for systems with few degrees of freedom. A proper thermostat should keep the average temperature constant and allow fluctuations of the instantaneous temperature with a standard deviation given by:

$$\sigma(T) = T_0 \sqrt{\frac{2}{N_f}} \quad (4)$$

In practice, thermostating is achieved through additional terms in the equation of motion, which effectively accelerate or slow the particles. The fundamental equation for this process is the Langevin equation:

$$\frac{d\mathbf{v}_A}{dt} = \frac{\mathbf{F}_A(t)}{m} - \gamma\mathbf{v}_A(t) + \mathbf{R}(t) \quad (5)$$

This equation extends Eq. (2) by two additional terms. The first describes the loss of kinetic energy to the environment due to friction. The second is a random process and represents collisions with particles in the environment. If γ and $\mathbf{R}(t)$ are chosen so that they obey the fluctuation-dissipation theorem, correct sampling of the canonical ensemble is achieved.

In EMPIRE, this is implemented as the LANGEVIN thermostat [39], where a time constant is provided so that

$$\gamma = \frac{1}{\tau} \quad (6)$$

and

$$\mathbf{R}(t) = \nu_A(t) \sqrt{\gamma(2-\gamma) \frac{k_B T_0}{m_A}} W \quad (7)$$

where W is a normal distributed random number.

It can be desirable to avoid the use of random numbers in the thermostat, in order to obtain deterministic trajectories. The simplest way of doing this is to rescale all velocities periodically by a factor proportional to the difference between the instantaneous and the desired temperature. In the extreme case, this principle can be used to fix the instantaneous temperature to the desired average temperature using:

$$\gamma = \sqrt{\frac{T_0}{T}} \quad (8)$$

As mentioned above, this is not useful for sampling the canonical ensemble, where fluctuations of the instantaneous temperature are expected (although it is quite useful for quickly thermalizing a system to a given temperature). Berendsen developed a damped version of the velocity-rescaling thermostat, where the instantaneous temperature is scaled to approach the desired average exponentially with a given time constant.

$$\lambda = \sqrt{1 - \frac{\Delta t}{\tau} \cdot \frac{(T - T_0)}{T}} \quad (9)$$

This algorithm is implemented in EMPIRE as the BERENDSEN thermostat [40]. Please note that if the time constant is chosen to be equal to the MD timestep, Eq. (9) reduces to Eq. (8).

Unlike the Langevin thermostat, the Berendsen thermostat does not strictly sample the canonical ensemble but provides a good approximation to it if a large enough relaxation time is chosen (typically between 100 and 400 fs). One consequence of this is the so-called flying ice cube effect [41], which causes energy to flow from high- to low-frequency degrees of freedom during the simulation, in violation of the equipartition principle. This is particularly noticeable because it also causes

energy transfer to the translation and rotational degrees of freedom (and therefore violates the conservation of linear and angular momentum). The latter is avoided in EMPIRE by periodically removing the overall translation and rotation, but more subtle equipartition artifacts will still be present in long simulations using the Berendsen thermostat.

The random collisions with virtual particles in the Langevin thermostat also cause translation and rotation of the system. However, this motion does not violate conservation of momentum of the overall system including the virtual particles. In fact, if the translational and rotational motion is removed, the dynamics no longer sample the canonical ensemble correctly. Typically, center-of-motion translation for Langevin trajectories is removed from the trajectory for analysis and visualization post hoc.

Technical aspects

SCF convergence and energy conservation

While integrating Eq. (2) with the Velocity-Verlet algorithm in principle should conserve the total energy, this is not necessarily the case in quantum mechanical dynamics, because the self-consistent field approach only converges the energy to a certain threshold. In EMPIRE, the default value is 10^{-4} kcal mol⁻¹, which is sufficient to conserve the energy if a new extended Hückel (EH) initial guess is made at each MD step. Using the wavefunction of the previous MD step as the initial guess can accelerate SCF convergence but in this case, the default convergence threshold is insufficient, since the converged wavefunction will be biased systematically towards the previous step. In practice, this means that NVE dynamics using the last wavefunction as initial guess loose energy proportionally to the SCF convergence criterion (see Fig. 1). This could be avoided using an extrapolation scheme to generate the initial guess based on two or more previous steps. However, since this increases the memory requirements (which are often the limiting factor in large simulations), we did not consider this practical.

When doing NVE dynamics, it is therefore recommended to recalculate the EH initial guess at each step (which is the default), since SCF convergence is typically very fast. Should SCF convergence be unusually slow in a specific system, using the previous wavefunction combined with a tight convergence criterion may lead to better performance. For NVT dynamics, the thermostat compensates the energy loss, and the default convergence criterion combined with the previous wavefunction as initial guess can therefore be used.

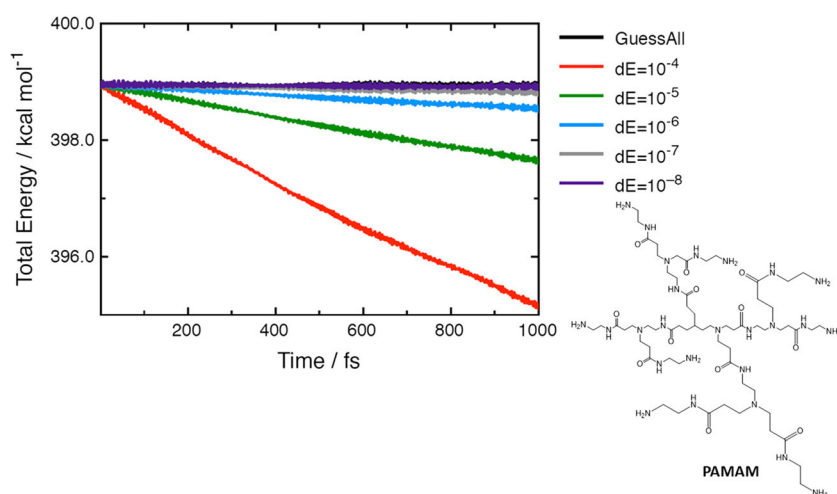


Fig. 1 Total energy conservation for a second-generation PAMAM dendrimer with different SCF convergence thresholds

Comparison of thermostats

As discussed above, the instantaneous temperature in the canonical ensemble should fluctuate according to Eq. (3). For the Langevin thermostat, this condition is fulfilled (by construction) for any thermostat time constant, although very short time constants do not provide smooth trajectories and effectively resemble Monte-Carlo simulations. To visualize this, Fig. 2 shows the temperature distributions for simulations using Langevin thermostats with time constants of 10 and 100 fs. Clearly, the time constant does not affect the statistical properties of the ensemble, although the actual trajectories are different.

This is different for the Berendsen thermostat, where the temperature distribution depends strongly on the time constant. For production calculations, it is important to choose a sufficiently large time constant to avoid an unrealistically strong interference of the thermostat with the dynamics. In fact, it is often recommended to use the thermostat only for equilibration and to calculate any properties using the microcanonical ensemble. However, in this case, conformational changes during the

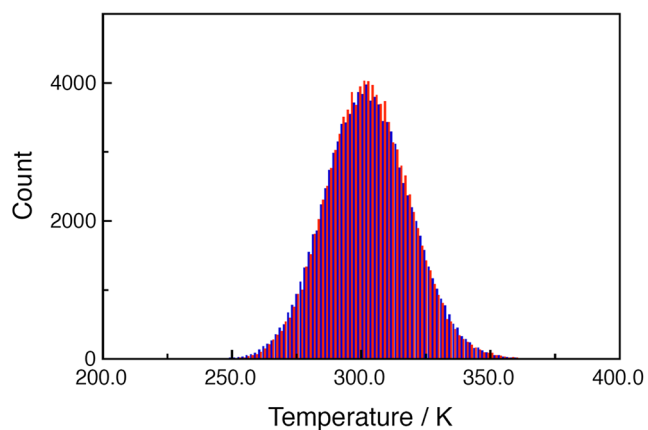


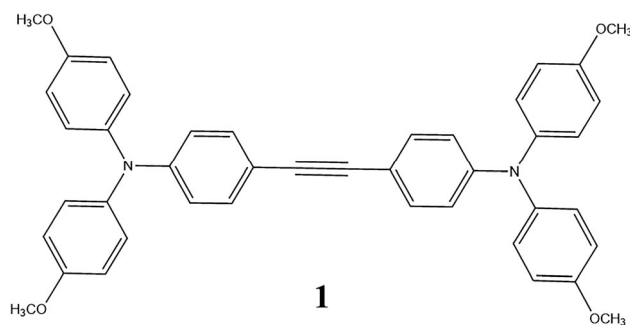
Fig. 2 Histogram for temperature distributions with Langevin thermostat time constants of 10 (blue) and 100 fs (red)

dynamics can lead to large changes in the average temperature and possibly render the simulation invalid. As mentioned above, time constants between 100 and 400 fs are typically appropriate.

Applications

Reorganization dynamics upon ionization

To demonstrate the use of direct semiempirical molecular dynamics for fast charge-transfer processes, the geometry of compound **1** has been investigated in detail by Lambert and Nöll [42].



The geometry of ground state **1** was first optimized using the AM1 Hamiltonian [8] and this geometry used as the starting point for UHF simulations of the radical cation within an NVE ensemble. The simulation was run for 20 ns with a timestep of 0.25 fs. Figure 3 shows a time-dependent trace of the calculated dipole moment (relative to the center of gravity of the cation) with plots of the molecular electrostatic potential [43] projected onto the 0.01 a.u. isodensity surface (all calculated with AM1). The calculation gives two alternative radical cation states localized on each of the triphenylamine moieties.

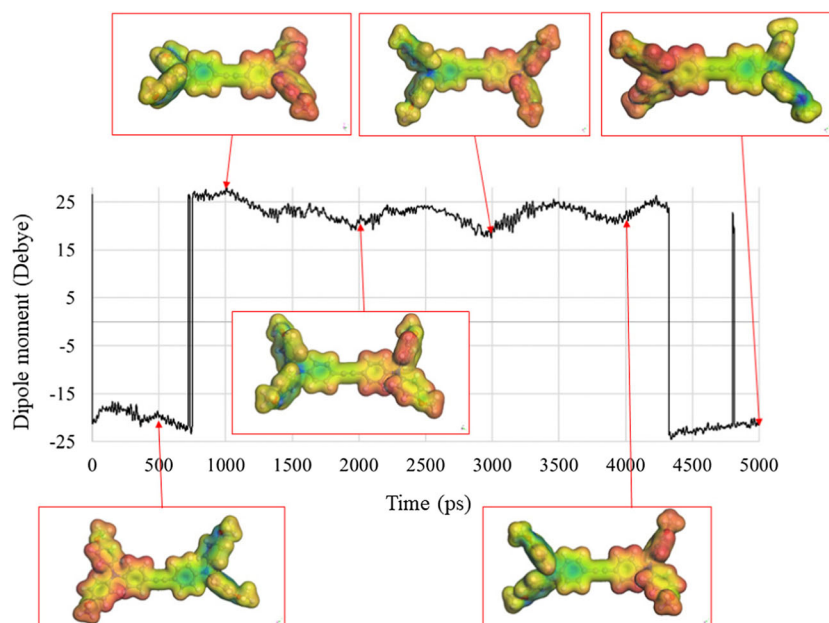


Fig. 3 Time dependence of the dipole moment along the molecular axis and snapshots from an NVE dynamic simulation of the radical cation of compound **1**. The UHF/AM1 calculated electrostatic potential projected

onto the 0.01 a.u. isodensity surface is shown on a scale of -70 (blue) to $+170$ (red) kcal mol $^{-1}$ for selected snapshots

The sign of the dipole moment indicates the oxidized triphenylamine moiety (negative for the left-hand side and positive for the right in the diagrams). Charge-transfer (CT) events from one side to the other are observed in the simulation with a longest lifetime for a single CT-state of approximately 3.5 ns. These events are accompanied by a breakdown of energy conservation in the NVE ensemble. The total energy increases by approximately 3 kcal mol $^{-1}$ at the first transition and 1.6 kcal mol $^{-1}$ at

the second, as shown in Figure S1 of the Supporting Information. The corresponding trace for the kinetic energy is shown in Figure S2.

Nanosecond dynamics of proton escape from NH $_4^+$ @C $_{60}$

MNDO-F [44, 45] was used to investigate the escape of a proton through the fullerene wall for the reaction

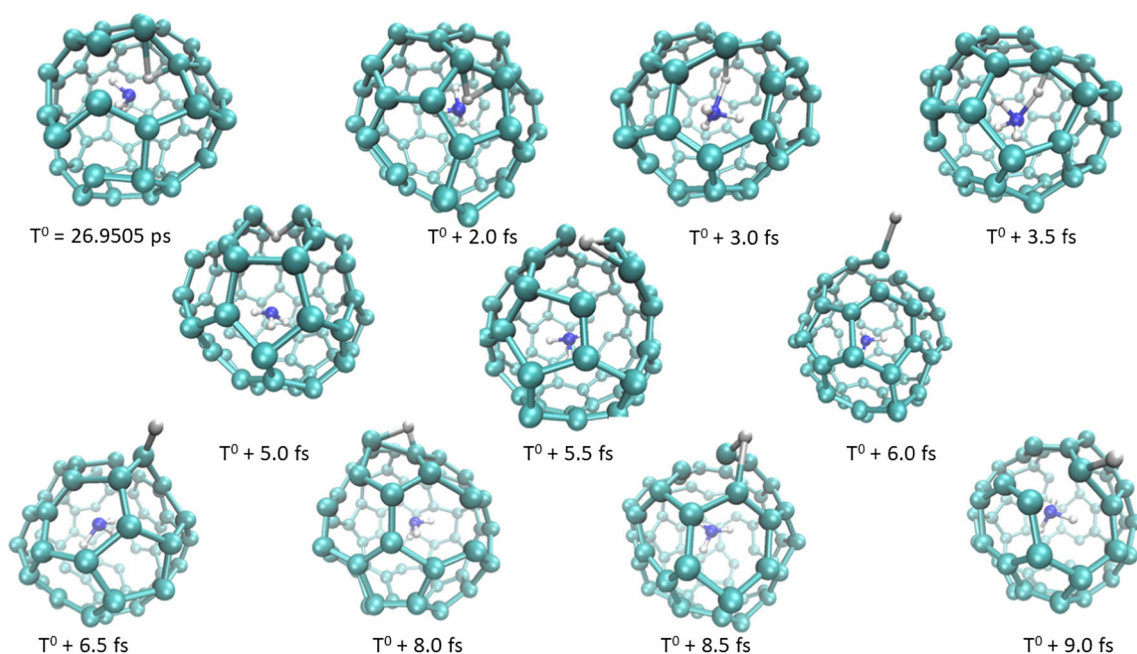


Fig. 4 Snapshots taken at the times indicated in the NH $_4^+$ @C $_{60}$ simulation at 4500 K as the proton leaves the fullerene cage

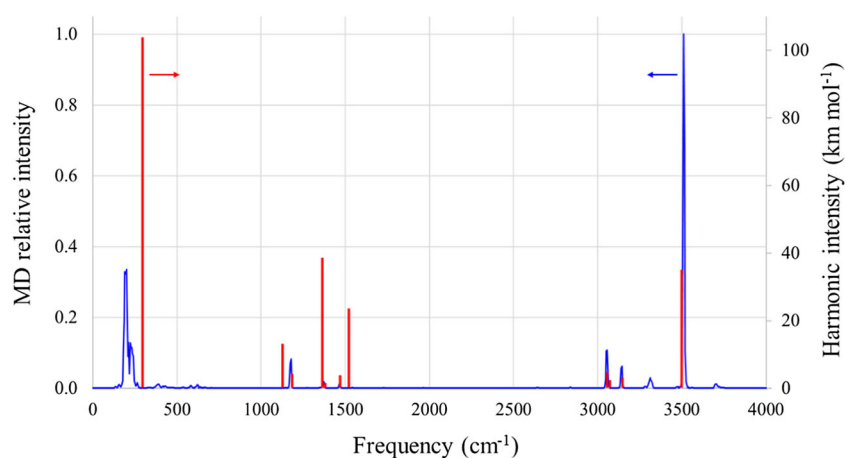


Fig. 5 MD-based AM1 IR spectrum (blue line) for methanol at 300 K. The red bars indicate the calculated AM1 harmonic frequencies and intensities

$\text{NH}_4^+@C_{60} \rightarrow \text{NH}_3@C_{60}\text{H}^+$ [46]. This reaction is calculated to be endothermic by $6.8 \text{ kcal mol}^{-1}$ at B3LYP/6-31G(d) [46] but exothermic by $9.6 \text{ kcal mol}^{-1}$ with MNDO-F, so that 1.5 ns NVT-simulations were started with the MNDO-F-optimized geometry of $\text{NH}_4^+@C_{60}$ at temperatures of 3000, 3500, and 4000 K. The simulations at the two lower temperatures did not result in escape of a proton but that at 4000 K gave a reaction after almost 27 ps, as shown in Fig. 4.

The escape reaction starts by forming an endohedral C–H bond at T^0 . Two femtoseconds later, an endohedrally proton-bridged C–C bond is observed, followed by structures at $T^0 + 3.0$ and $T^0 + 3.5$ femtoseconds in which the proton is shared between N and a carbon atom of the cage. The actual escape event starts at $T^0 + 5.0$ femtoseconds with the proton bridging an opened cage C–C bond endohedrally, so that it can swing through the fullerene wall ($T^0 + 5.5$) to give a series of exohedrally protonated $\text{NH}_3@C_{60}$ structures with partially

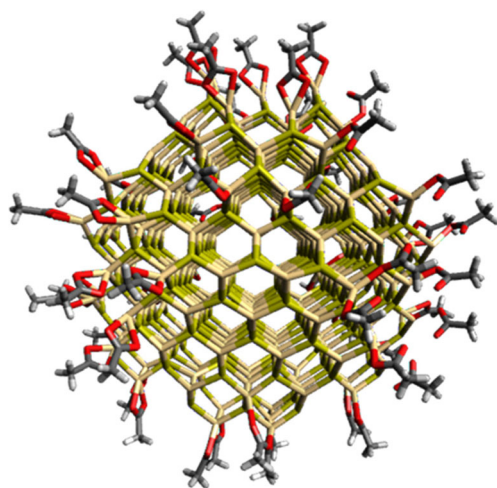


Fig. 6 MNDO/D-equilibrated geometry of an acetate-capped CdS quantum dot at 300 K

opened cages ($T^0 + 6.0$ to $T^0 + 9.0$). These structures persist for the remainder of the simulation.

Vibrational spectrum of a CdS quantum dot

BOMD trajectories can be used to simulate vibrational spectra. For example, infrared (IR) absorption and transmission spectra can be simulated via the Fourier transformation of the dipole moment autocorrelation function. This has several advantages compared with the more established route in the context of a “static” normal-mode calculation within the harmonic approximation. Most importantly, the full anharmonicity of the potential energy surface is taken into account in this way. Furthermore, a full spectrum is obtained, meaning that no assumptions about broadening and line shapes are made. This also allows the investigation of thermal effects.

We follow the protocol of Kirchner and co-workers [47] for simulating vibrational spectra from EMPIRE MD trajectories. Briefly, this entails first (numerically) calculating the time derivative of the dipole vector. Then, the autocorrelation function of the dipole derivative is computed. Finally, this autocorrelation function is Fourier-transformed to obtain the IR spectrum.

Figure 5 shows the results obtained for methanol (AM1 Hamiltonian, 2^{14} 0.5 fs MD steps, NVT ensemble at 300 K).

For comparison, the normal-mode frequencies are shown as red bars, as computed from a conventional static frequency calculation with AM1 within the harmonic approximation using Gaussian 16 [48]. It can be seen that the MD-derived peaks correlate closely with the location of the normal modes. Here, the largest deviation is observed for the H–C–O–H torsional mode below 300 cm^{-1} , which is caused by anharmonic effects [49]. In contrast, the relative intensities are only comparable with the harmonic calculation in the intermediate region between 1000 and 1500 cm^{-1} (see Table S1 of the ESI). Most prominently, the OH-stretching band around 3500 cm^{-1} is predicted to be the most intense in the MD calculation,

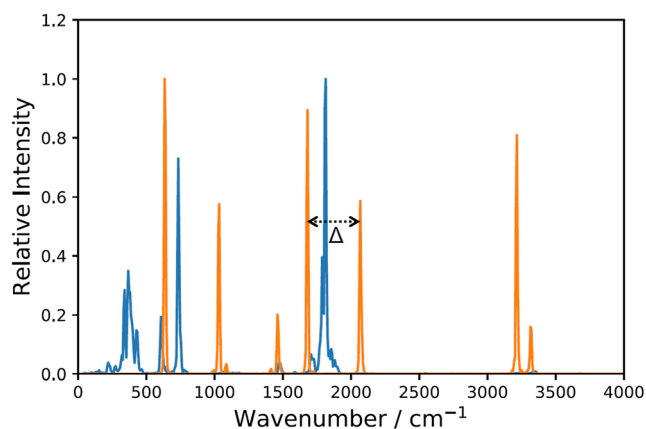


Fig. 7 MD-based IR spectrum for an acetate-capped CdS quantum dot (shown in the inset) at 300 K (blue). For comparison, the spectrum of isolated acetate is shown in orange. All simulations used the MNDO/D Hamiltonian within the NVT ensemble at 300 K with 0.5 fs time steps. Ten thousand timesteps were analyzed after 2.000 steps of equilibration

while it is much weaker in the harmonic approximation. Importantly, the MD-based simulation is in much better agreement with experiment in this case, where this band is also the most intense. Indeed, the same phenomenon has been observed for DFT-based calculations; it was ascribed to the prominent role of anharmonic effects for this band [47].

The big advantage of semiempirical over DFT-based MD simulations lies in the accessible timescales and sizes. To illustrate this point, the MD-derived IR spectrum of the acetate-capped CdS quantum dot shown in Fig. 6 and consisting of 836 atoms is shown in Fig. 7.

This simulation required on average 9 s per timestep on 20 cores, making it routinely feasible even with relatively modest resources. The MNDO/d Hamiltonian [50], which uses the original MNDO parameters for first row atoms, was used for this simulation. While the spectrum of isolated acetate shows significant shifts with respect to the experimental band positions (mainly due to the limitations of minimal basis sets for an anionic systems), the effect of coordination on individual bands can be well studied. For example, the symmetric and asymmetric COO-stretching vibrations are shifted to considerably higher frequencies (ca. 400 cm^{-1} relative to PBE/mTzVP) but the difference between the two frequencies (Δ) is in good agreement (ca. 350 cm^{-1}). Upon coordination to the quantum dot, this difference vanishes and a single broad signal is observed. This is due to the predominantly bidentate ligation to surface Cd atoms. The same trend (small or vanishing Δ) has been reported experimentally. [51, 52]

Conclusions

NDDO-based semiempirical MO theory is well suited for Born-Oppenheimer MD simulations on systems of several thousand atoms using parallel hardware and software.

Single-node performance of six timesteps per minute for 1000 atoms is easily achievable. Similar turnaround times are realistic on multi-node clusters for larger systems. The advantages of direct quantum mechanical MD therefore now become available for large molecules and aggregates, including those for which linear-scaling approaches through local approximations are not applicable.

Funding information Open Access funding provided by Projekt DEAL. This work was financially supported by the Bavarian Government within the “Solar Technologies go Hybrid” (SolTech) initiative.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Margraf JT, Hennemann M, Meyer B, Clark T (2015) EMPIRE: a highly parallel semiempirical molecular orbital program: 2: periodic boundary conditions. *J Mol Model* 21:1. <https://doi.org/10.1007/s00894-015-2692-3>
- Hennemann M, Clark T (2014) EMPIRE: a highly parallel semiempirical molecular orbital program: 1: self-consistent field calculations. *J Mol Model* 20:2331. <https://doi.org/10.1007/s00894-014-2331-4>
- Pople JA, Beveridge DL, Dobosh PA (1967) Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. *J Chem Phys* 47:2026–2033. <https://doi.org/10.1063/1.1712233>
- Pople JA, Santry DP, Segal GA (1965) Approximate self-consistent molecular orbital theory. I. Invariant procedures. *J Chem Phys* 43: S129. <https://doi.org/10.1063/1.1701475>
- Dewar MJS, Thiel W (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J Am Chem Soc* 99:4899–4907. <https://doi.org/10.1021/ja00457a004>
- Dewar MJS, Thiel W (1977) Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J Am Chem Soc* 99:4907–4917. <https://doi.org/10.1021/ja00457a005>
- Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 107:3902–3909. <https://doi.org/10.1021/ja00299a024>
- Stewart JJP (2004) Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *J Mol Model* 10:155–164. <https://doi.org/10.1007/s00894-004-0183-z>
- Stewart JJP (1989) *J Comput Chem* 10:209. *J Comput Chem* 10: 209–220. <https://doi.org/10.1002/jcc.540100208>

10. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* 13:1173. <https://doi.org/10.1007/s00894-007-0233-4>
11. Thiel W (2014) Semiempirical quantum-chemical methods. *WIREs Comput Mol Sci* 4(2014):145–157. <https://doi.org/10.1002/wcms.1161>
12. Thiel W, Voityuk A (1996) Extension of MNDO to d orbitals: parameters and results for the second-row elements and for the zinc group. *J Phys Chem* 100:616. <https://doi.org/10.1021/jp952148o>
13. Akimov AV, Prezhdo OV (2015) Large-scale computations in chemistry: a bird's eye view of a vibrant field. *Chem Rev* 115: 5797–5890. <https://doi.org/10.1021/cr500524c>
14. Marx D, Hutter J (2009) *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, Cambridge ISBN-13: 978-0521898638
15. Oliveira AF, Seifert G, Heine T, Duarte HA (2009) Density-functional based tight-binding: an approximate DFT method. *J Braz Chem Soc* 20:1193–1205. <https://doi.org/10.1590/S0103-50532009000700002>
16. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of reaction of lysozyme. *J Mol Biol* 103:227–249. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9)
17. Field MJ, Bash PA, Karplus M (1990) A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J Comput Chem* 11:700–733. <https://doi.org/10.1002/jcc.540110605>
18. Walker RC, Crowley MF, Case DA (2008) The implementation of a fast and accurate QM/MM potential method in Amber. *J Comput Chem* 29:1019–1031. <https://doi.org/10.1002/jcc.20857>
19. Korth M, Pitoňák M, Rezáč J, Hobza P (2010) A transferable H-bonding correction for semiempirical quantum-chemical methods. *J Chem Theory Comp* 6:344–352. <https://doi.org/10.1021/ct900541n>
20. Rezáč J, Fanfrlík J, Salahub D, Hobza P (2009) Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *J Chem Theory Comp* 5:1749–1760. <https://doi.org/10.1021/ct9000922>
21. Dral PO, Wu X, Spörkel L, Kosłowski A, Thiel W (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: benchmarks for ground-state properties. *J Chem Theory Comput* 12:1097–1120. <https://doi.org/10.1021/acs.jctc.5b01047>
22. Ryan H, Carter M, Stenmark P, Stewart JJP, Braun-Sand SB (2016) A comparison of X-ray and calculated structures of the enzyme MTH1. *J Mol Model* 22:168. <https://doi.org/10.1007/s00894-016-3025-x>
23. Tuttle T, Thiel W (2008) OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys Chem Chem Phys* 10:2159–2166. <https://doi.org/10.1039/B718795E>
24. Rezáč J, Hobza P (2012) Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J Chem Theory Comput* 8:141–151. <https://doi.org/10.1021/ct200751e>
25. Rezáč J, Hobza P (2011) A halogen-bonding correction for the semiempirical PM6 method. *Chem Phys Lett* 506:286–289. <https://doi.org/10.1016/j.cplett.2011.03.009>
26. Bauer CA, Grimme S (2016) How to compute electron ionization mass spectra from first principles. *J Phys Chem A* 120:3755–3766. <https://doi.org/10.1021/acs.jpca.6b02907>
27. McClellan JJ, Hughes TF, Bartlett RJ (2005) Application of the transfer Hamiltonian formalism to high-energy model systems. *Int J Quantum Chem* 105:914–920. <https://doi.org/10.1002/qua.20685>
28. Bartlett RJ, De Taylor C, Korkin A (2005) Achieving predictive simulations with quantum mechanical forces via the transfer hamiltonian: problems and prospects. In: Yip S (ed) *Handbook of materials modeling SE - 4*. Springer, Dordrecht, pp 27–57. https://doi.org/10.1007/978-1-4020-3286-8_4
29. Taylor CE, Cory MG, Bartlett RJ, Thiel W (2003) The transfer Hamiltonian: a tool for large scale simulations with quantum mechanical forces. *Comput Mater Sci* 27:204–211. [https://doi.org/10.1016/S0927-0256\(03\)00002-8](https://doi.org/10.1016/S0927-0256(03)00002-8)
30. Taylor DE, Runge K, Bartlett RJ (2005) Study of the effect of hydration on the tensile strength of a silica nanotube. *Mol Phys* 103:2019–2026. <https://doi.org/10.1080/00268970500131199>
31. Spörkel L, Thiel W (2016) Adaptive time steps in trajectory surface hopping simulations. *J Chem Phys* 144:194108. <https://doi.org/10.1063/1.4948956>
32. Pal S, Trivedi DJ, Akimov AV, Aradi B, Frauenheim T, Prezhdo OV (2016) Nonadiabatic molecular dynamics for thousand atom systems: a tight-binding approach toward PYXAID. *J Chem Theory Comput* 12:1436–1448. <https://doi.org/10.1021/acs.jctc.5b01231>
33. Stewart JJP (1996) Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem* 58:133–146. [https://doi.org/10.1002/\(SICI\)1097-461X\(1996\)58:2<133::AID-QUA2>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-461X(1996)58:2<133::AID-QUA2>3.0.CO;2-Z)
34. Dixon SL, Merz Jr KM (1997) Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J Chem Phys* 107: 879–893. <https://doi.org/10.1063/1.474386>
35. Wick CR, Hennemann M, Stewart JJP, Clark T (2014) Self-consistent field convergence for proteins: a comparison of full and localized-molecular-orbital schemes. *J Mol Model* 20:2159. <https://doi.org/10.1007/s00894-014-2159-y>
36. Born M, Oppenheimer R (1927) Zur Quantentheorie der Molekeln. *Ann Phys* 389:457–484. <https://doi.org/10.1002/andp.19273892002>
37. Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett* 55:2471–2474. <https://doi.org/10.1103/PhysRevLett.55.2471>
38. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J Chem Phys* 76:637–649. <https://doi.org/10.1063/1.442716>
39. Soddemann T, Dünweg B, Kremer K (2003) Dissipative particle dynamics: a useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. *Phys Rev E* 68:046702. <https://doi.org/10.1103/PhysRevE.68.046702>
40. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690. <https://doi.org/10.1063/1.448118>
41. Harvey SC, Tan RK-Z, Cheatham TE (1998) The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J Comput Chem* 19:726–740. [https://doi.org/10.1002/\(SICI\)1096-987X\(199805\)19:7<726::AID-JCC4>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-987X(199805)19:7<726::AID-JCC4>3.0.CO;2-S)
42. Lambert C, Nöll G (2002, 2039–2043) Tuning of intervalence charge transfer energies by substituents in one-dimensional bis(triarylamine) systems. *J Chem Soc Perkin Trans II*. <https://doi.org/10.1039/b207736a>
43. Horn AHC, Lin J-H, Clark T (2005) Multipole electrostatic model for MNDO-like techniques with minimal valence *spd*-basis sets. *Theor Chem Acc* 114:159–168. <https://doi.org/10.1007/s00214-005-0657-9>, Erratum (2007) *Theor Chem Acc* 117:461–465. <https://doi.org/10.1007/s00214-006-0167-4>
44. Kriebel M, Weber K, Clark T (2018) A Feynman dispersion correction: a proof of principle for MNDO. *J Mol Model* 24:338. <https://doi.org/10.1007/s00894-018-3874-6>
45. Kriebel M, Heßelmann A, Hennemann M, Clark T (2019) The Feynman dispersion correction for MNDO extended to F, Cl, Br and I. *J Mol Model* 25: DOI: <https://doi.org/10.1007/s00894-019-4038-z> : (2019)

- Erratum. *J Mol Model* 25:257. <https://doi.org/10.1007/s00894-019-4142-0>
46. Dral PO, Clark T (2017) On the feasibility of reactions through the fullerene wall: a theoretical study of $\text{NH}_x@C_{60}$. *Phys Chem Chem Phys* 19:17199–17209. <https://doi.org/10.1039/C7CP02865B>
47. Thomas M, Brehm M, Fligg R, Vöhringer P, Kirchner B (2013) Computing vibrational spectra from ab initio molecular dynamics. *Phys Chem Chem Phys* 15:6608–6622. <https://doi.org/10.1039/C3CP44302G>
48. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Camm, R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian 16, Revision A.03, Gaussian, Inc., Wallingford CT. <https://gaussian.com/gaussian16/> accessed 25th September 2019
49. Bowman JM (1986) The self-consistent-field approach to polyatomic vibrations. *Acc Chem Res* 19:202–208. <https://doi.org/10.1021/ar00127a002>
50. Thiel W, Voityuk AA (1996) Extension of MNDO to d orbitals: parameters and results for the second-row elements and for the zinc group. *J Phys Chem* 100:616–626. <https://doi.org/10.1021/jp952148o>
51. Deacon BG, Phillips RJ (1980) Relationships between the carbon-oxygen stretching frequencies of carboxylate complexes and the type of carboxylate coordination. *Coord Chem Rev* 33:227–250. [https://doi.org/10.1016/S0010-8545\(00\)80455-5](https://doi.org/10.1016/S0010-8545(00)80455-5)
52. Margraf JT, Ruland A, Sgobba V, Guldi DM, Clark T (2013) Theoretical and experimental insights into the surface chemistry of semiconductor quantum dots. *Langmuir* 29:15450–15456. <https://doi.org/10.1021/la403633e>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.